

Vergleich des SIFT- und SURF-Algorithmus

Essay

1. Abstract

SIFT- und SURF-Algorithmus sind sich in der Grundidee zur Objekterkennung in Bildern sehr ähnlich. Beide Algorithmen klassifizieren Objekte im Bild, indem sie deren Intensitätswerte als Funktionen approximieren. Die Eigenschaften dieser Funktionen werden durch Ableiten untersucht, sodass daraus möglichst viel Information extrahiert wird. Wie ein Fingerabdruck kann die Information gespeichert und mit anderen „Fingerabdrücken“ verglichen werden, um gemeinsame Merkmale in verschiedenen Bildern zu finden.

Bei der Untersuchung der beiden Algorithmen hat sich ergeben, dass sich die Algorithmen in Genauigkeit und Effizienz bei der Objekterkennung unterscheiden. Denn während der SURF-Algorithmus, auf Effizienz bedacht, grobe Approximationen bezüglich der mathematischen Berechnungen trifft, wird im SIFT-Algorithmus nahezu jedes einzelne Pixel und dessen Umgebung sehr genau betrachtet, was einen verhältnismäßig hohen Rechenaufwand erfordert.

2. Einführung

Effiziente und robuste Algorithmen zur Verarbeitung von Bildern und Bildsequenzen erfahren durch die Automatisierung in der Robotik und der Automobilindustrie sowie in anderen Bereichen unserer digitalisierten Umwelt einen starken Focus.

In diesem Essay werden der SIFT- und der SURF-Algorithmus verglichen. Beide Algorithmen dienen der Klassifizierung und Wiedererkennung von Objekten in Bildern.

Bei der Untersuchung der Algorithmen werden in diesem Essay Aussagen bezüglich der Robustheit und Effizienz getroffen. Zudem wird eine Empfehlung hinsichtlich der Anwendungsgebiete der Algorithmen gegeben.

3. Hauptteil

Allgemein genügt die Objekterkennung in einem Bild folgenden 3 Schritten [1]:

- Selektion von geeigneten Kandidaten für eine markante Charakterisierung
- Beschreibung der Umgebung der jeweiligen selektierten Punkte im Bild mittels *Deskriptoren*
- Vergleich der Deskriptoren mit Deskriptoren aus anderen Bilddaten, die beispielsweise durch Aufnahmen aus verschiedenen Blickwinkeln generiert wurden

3.1. Selektion geeigneter Kandidaten mittels SIFT

Im SIFT- sowie SURF-Algorithmus wird entsprechend des ersten Schrittes zunächst der 3-dimensionale *Skalenraum* $L(x,y,\sigma)$ betrachtet:

$$L(x, y, \sigma) = (G(x, y, \sigma)) * I(x, y) \quad (3.1)$$

Die diskreten Pixelwerte $I(x,y)$ des Bildes werden hierzu mit einem Gaußfilter mit variablem σ geglättet.

Der SIFT-Algorithmus berechnet infolge geeigneter Abtastfrequenz die sog. „*Difference of Gaussians*“, wobei der Skalierungswert $k=\frac{1}{\sqrt{2}}$ gewählt sei [1]:

$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) \quad (3.2)$$

Geeignete Kandidaten werden eruiert, indem jedes abgetastete Pixel mit seinen insgesamt 26 benachbarten Pixeln im durch D definierten Raum verglichen wird. Maxima und Minima werden als geeignete Kandidaten registriert.

Um Extremwerte, die beispielsweise durch Rauschen detektiert wurden, in weitere Berechnungen nicht einzubeziehen, erfolgt basierend auf der folgenden Formel eine weitere Selektion:

$$D(\hat{x}, \hat{y}, \hat{\sigma}) = D + \frac{1}{2} \frac{\delta D}{\delta(x, y, \sigma)} \hat{x}, \hat{y}, \hat{\sigma} \quad (3.3)$$

mit dem Offset:

$$\hat{x} = x - x_s \quad (3.4)$$

$$\hat{y} = y - y_s \quad (3.5)$$

$$\hat{\sigma} = \sigma - \sigma_s \quad (3.6)$$

$D(x,y,\sigma)$ lässt sich durch die Taylorentwicklung um den jeweiligen Samplevektor (x_s, y_s, σ_s) ableiten, um letztendlich Gleichung (3.3) zu erhalten [1]. Daraufhin werden alle Extreme herausgefiltert, die folgendem Kriterium nicht genügen [1]:

$$|D(\hat{x}, \hat{y}, \hat{\sigma})| > 0.03 \quad (3.7)$$

3.2. Aufwand und Robustheit der Selektion mittels SIFT

Der Aufwand der Faltung in Gleichung (3.1) hängt von der Anzahl der Pixel im Bild ab, da auf jedes Pixel der Gaußfilter angewandt wird. Für ein Bild der Auflösung $(n \times m)$ werden folglich $n \times m$ Iterationen benötigt. Somit ist die Berechnung nach Gleichung (3.1) relativ aufwendig.

Die Glättung durch Faltung ist aufgrund der Berücksichtigung aller Pixel jedoch sehr robust.

Entsprechend Gleichung (3.2) wird die Differenz aus zwei Glättungen mit verschieden skaliertem σ berechnet. Somit erhöht sich der Rechenaufwand lediglich um eine Differenz pro Pixel und ist vernachlässigbar gering.

Der Aufwand der Taylorentwicklungen aus Gleichung (3.3) um die jeweiligen geeigneten Kandidaten hängt direkt von der Anzahl geeigneter Kandidaten ab. Denn die Taylorreihe muss um jeden Kandidaten neu entwickelt werden, um durch Rauschen erzeugte Kandidaten abzugrenzen. Dies ist meiner Meinung relativ aufwendig und hat negative Auswirkungen auf die Robustheit, da eine 3-dimensionale Ableitung um jeden Entwicklungspunkt gebildet werden muss, was durchaus zu numerischen Instabilitäten führen kann. Insgesamt erfordert die Verarbeitung der Pixel mit Gleichungen (3.1-7) im SIFT-Algorithmus folglich einen hohen Rechenaufwand.

3.3. Selektion geeigneter Kandidaten mittels SURF

Im SURF-Algorithmus werden mittels Approximation der Hessematrix geeignete Kandidaten selektiert. Die Hessematrix im Skalenraum ist wie folgt definiert:

$$H(x, y, \sigma) = \begin{bmatrix} L_{xx}(x, y, \sigma) & L_{xy}(x, y, \sigma) \\ L_{xy}(x, y, \sigma) & L_{yy}(x, y, \sigma) \end{bmatrix} \quad (3.8)$$

Durch Berechnung der Maxima der Determinante werden geeignete Kandidaten eruiert.

Die Hessematrix wird nach Gleichung (3.8) zur Ermittlung geeigneter Kandidaten aus Effizienzgründen mit sog. *Box-Filtern* unter Verwendung von integralen Bildern approximiert. Somit lassen sich geeignete Kandidaten über die approximierte Hessematrix über die Determinante berechnen:

$$\det(H_{approx}) = D_{xx} D_{yy} - (w D_{xy})^2 \quad (3.9)$$

Die Eintrag D_{xx} ist eine Approximation von L_{xx} . Analoges gilt für D_{yy} und D_{xy} . Zudem gewichtet der SURF-Algorithmus D_{xy} mit dem konstanten Faktor $w=0.9$ [2].

3.4. Aufwand und Robustheit der Selektion mittels SURF

Die Verwendung von integralen Bildern ist sehr effizient, da der Aufwand nicht von der Integralgröße abhängt [2]. Allerdings kann diese Approximation meiner Meinung nach das Rauschen erhöhen und führt zu Ungenauigkeiten bei der Detektion von geeigneten Kandidaten. Die Untersuchung der Maxima der Determinante nach Gleichung (3.9) hängt wiederum von der Anzahl geeigneter Kandidaten ab. Meiner Meinung nach sollte im Algorithmus entsprechend Gleichung (3.7) noch ein Schwellwert eingeführt werden, um durch Rauschen erzeugte Kandidaten zu reduzieren. Dies würde letztendlich zur Robustheit des Algorithmus beitragen.

3.5. Klassifizierung geeigneter Kandidaten mittels SIFT und SURF

Im SIFT – Algorithmus wird folgende Formeln verwendet, um jedem gesampelten Pixel $L(x,y)$ einen Gradienten sowie eine Orientierung zuzuweisen:

$$m(x, y) = \sqrt{(L(x + 1, y) - L(x - y, y))^2 + (L(x, y + 1) - L(x, y - 1))^2} \quad (3.10)$$

$$\theta(x,y) = \tan^{-1}\left(\frac{L(x,y+1) - L(x,y-1)}{L(x+1,y) - L(x-1,y)}\right) \quad (3.11)$$

Daraufhin wird ein Histogramm generiert, welches die Orientierung und die Gradienten in einer Umgebung eines jeden selektierten Kandidaten speichert.

Im SURF – Algorithmus wird die Verteilung der Intensität $I(x,y)$ der selektierten Pixel sowie die Verteilung der Intensität um die selektierte Pixel herum mittels „Haar Wavelet“ approximiert. Somit erhält man die Steigung und Varianz der Intensität jedes selektierten Kandidaten. Basierend auf der Steigung und Varianz in einer Umgebung um den selektierten Kandidaten werden die sog. *SURF-Merkmalvektoren* erzeugt und mit Merkmalsvektoren aus anderen Bildern abgeglichen.

3.5. Bewertung des Aufwands zur Klassifizierung

Jedem einzelnen Pixel im Bild wird entsprechend der Gleichungen (3.10-11) ein Gradient und eine Orientierung zugeordnet. Die Berechnungen nach Gleichung (3.10-11) können parallel durchgeführt werden und der Rechenaufwand ist von der Anzahl der Pixel abhängig.

Der zweite Schritt, die Umgebung jedes selektierten Kandidaten in einem Histogramm abzuspeichern ist vom Radius der Umgebung sowie der Anzahl selektierten Kandidaten abhängig. Insgesamt lässt sich also folgern, dass der Aufwand zur Beschreibung von Objekten im Bild mittels Merkmalsvektoren im SIFT-Algorithmus erheblich ist, da bei entsprechender Abtastfrequenz auf nahezu jedes Pixel mehrere Rechenoperationen angewandt werden und bei entsprechender Genauigkeit die Dimension des Merkmalsvektors sehr groß werden kann.

Der SURF-Algorithmus verwendet zwar auch die Verteilung der Gradienten und die Orientierung für die Charakterisierung der Merkmalsvektoren, jedoch werden diese effizient mittels integrelem Filter approximiert. Diese Approximation verringert freilich die Genauigkeit.

4. Zusammenfassung

Als Fazit lässt sich festlegen, dass je nach Anwendung der passende Algorithmus für die Objekterkennung verwendet werden sollte. Eine meiner Meinung nach interessante Anwendung ist die Verarbeitung von Bildsequenzen in Echtzeit. Eine App, die infolge Echtzeitaufnahmen Merkmale im Gesicht derartig genau klassifiziert, dass man beispielsweise über eine Verknüpfung mit Facebook sofort den Namen und die Interessen des gefilmten Menschen erkennt.

Aufgrund begrenzter Rechenleistung (und Datenübertragung) sollte man sich folglich zunächst auf eine App mit implementierten SURF-Algorithmus beschränken. Dies könnte beispielsweise für die Objekterkennung von exotischen Tieren im Zoo geeignet sein.

Mit der Entwicklung der Technik sollte bei maximaler Rechenleistung und Datenübertragung bei ungedrosseltem Datenvolumen auch die Objekterkennung von Menschen mittels im Handy implementierten SIFT-Algorithmus in Echtzeit möglich sein.

Referenzen

- [1] David G.Lowe: *Distinctive Image Features from Scale-Invariant Keypoints*
- [2] Herbert Bay, Andreas Ess, Tinne Tuytelaars, Luc Van Gool: *Speeded-Up Robust Features (SURF)*